NTNU – Trondheim
Norwegian University of Science and Technology

Propensity score analysis:
Adjusting for multiple confounders in observational studies.

1 February 2013

Stian Lydersen

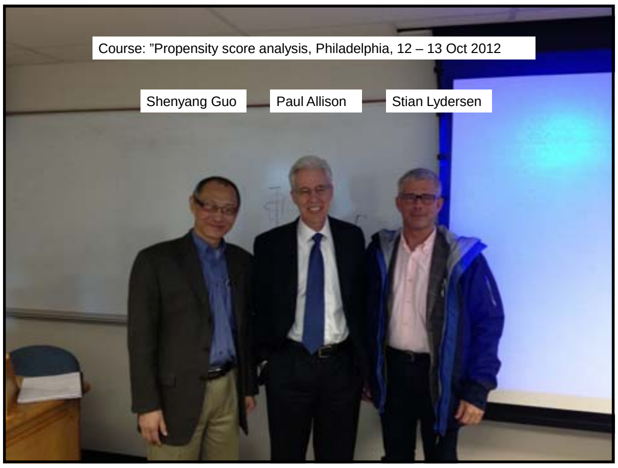
Version 1 February 2013

Propensity score analysis: Adjusting for multiple confounders in observational studies.

- Motivating example
- What is a confounder?
- How to adjust for confounders
- What is a propensity score?
- Estimating the propensity score
- Analysis adjusting for propensity score
- Example

Literature

- Katz, M. H. 2010, *Evaluating clinical and public health interventions a practical guide to study design and statistics* New York, Cambridge.
- Guo, S. & Fraser, M. W. 2010, *Propensity score analysis statistical methods and applications* Sage Publications, Thousand Oaks, Calif.
- D'Agostino, R. B. 1998, "Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group", *Statistics in Medicine*, vol. 17, no. 19, pp. 2265-2281.
- Rosenbaum, P. R. & Rubin, D. B. 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, vol. 70, no. 1, pp. 41-55.



Eur Child Adolesc Psychiatry
DOI 10.1007/s00787-012-0300-y

ORIGINAL CONTRIBUTION

Smoking during pregnancy and psychiatric disorders in preschoolers

Lise Carol Ellis · Turid Suzanne Berg-Nielsen · Stian Lydersen · Lars Wichstrøm

Received: 20 June 2011 / Accepted: 18 June 2012

Abstract The overall objective of this study was to determine whether smoking during pregnancy is related to psychiatric disorders in 4-year-olds while controlling for a wide range of potential confounding variables (i.e. parental anxiety, depression, personality disorders, drug abuse, and socio-economic characteristics). Parents of a community sample of 4-year-olds (N = 995) residing in the city of Trondheim, Norway were interviewed using the Preschool Age Psychiatric Assessment, which includes information on prenatal smoking. After adjusting for potential confounding variables using the propensity score, smoking during pregnancy was found to increase the odds for attention-deficit/hyperactivity disorder (ADHD) OR = 2.59 (CI 1.5–4.34, p < 0.001), oppositional defiant disorder (ODD) OR = 2.69 (CI 1.84–3.91, p = 0.02) and comorbid OR = 2.55 (CI 1.24–5.23, p < 0.001). Prenatal smoking during pregnancy is associated with an increased risk for symptoms of ADHD and ODD independently of each other, in 4-year-olds.

Keywords Prenatal smoking ADHD ODD Internalising disorders Preschool children

Introduction

Prenatal smoking has been found to increase the risk of attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD), poor cognitive functioning, antisocial problems, aggression, delinquency, substance abuse, and internalising problems [1–7]. The majority of studies on this topic have examined children during mid or late childhood. Children with an early manifestation of disruptive behaviours have been found to develop more serious long-term psychopathologies. For instance, approximately one-quarter of children with ODD later develop conduct disorder (CD), and a few of these children develop antisocial personality disorder in adulthood [8, 9]. The short- and long-term costs of these problems are grave not only for the patients and their families, but also for society at large. Therefore, it is important to establish whether prenatal smoking affects early development. At present, only seven studies have examined the effect of prenatal smoking on preschoolers [5, 10–15]. Six of these studies used various rating scales to measure the effects of prenatal smoking on symptoms

Ellis et al (2012): Smoking during pregnancy and psychiatric disorders in preschoolers.
Eur Child Adolesc Psychiatry

- Project "Tidlig trygg i Trondheim".
- Community sample of 995 pregnancies.
- Weighted sampling in four (low risk to high risk) groups based on the SDQ (Strength and difficulties) questionnaire. Sampling probabilities 0.37, 0.48, 0.70, 0.89.
- Exposure: Smoking during pregnancy (148 cases)
- Logistic regression with outcomes (events) at 4 years:
  - ADHD (attentiondeficit/hyperactivity disorder), 34 cases
  - ODD (oppositional defiant disorder), 57 cases
- Confounders:
  - Mother's age
  - SES (Socio-economic statius)
  - Antisocial personality traits
  - Plus 24 potential confounders

7

24 potential confounders:

- narcissistic personality traits
- histrionic personality traits
- borderline personality traits
- schizotypal personality traits
- paranoid personality traits
- avoidant personality traits
- dependent personality traits
- OCD personality traits
- parental alcohol use
- parental anxiety
- parental depresion
- alcohol use during pregnancy
- stress during pregnancy
- depression during pregnancy
- planned pregnancy
- parental feelings about pregnancy
- mothers' feelings in the first month after birth
- parental experience of mental breakdown
- parent requested medical treatment
- parent ever been arrested
- parent ever been indicted by police
- parental ability to pay family expenses
- parent received medical treatment for psychological disorder
- parental admission to a mental health institution

NTNU – Trondheim Norwegian University of Science and Technology

www.ntnu.edu

8

Propensity score analysis: Adjusting for multiple confounders in observational studies.

In observational studies (unlike in randomized controlled studies), the exposure (or treatment) groups are typically not balanced with respect to potential confounders. This is commonly handled by including the confounders as covariates in regression analyses. In studies with rare events, for example in logistic regression or Cox regression, this may not be possible due to many confounders compared to few cases with the event. This problem may be addressed using propensity score analysis.

NTNU – Trondheim Norwegian University of Science and Technology

www.ntnu.edu

9

Confounders:

- **What** is a confounder?
- **Why** adjust for confounders? Answer: Else, we introduce bias.
- **How** to adjust for confounders?

NTNU – Trondheim Norwegian University of Science and Technology

www.ntnu.edu

10

Definition of a confounder
(Rothman: "Epidemiology: An Introduction". 2nd ed. Oxford University Press, 2012, page 141.)

Confounding can be thought of as a mixing of effects. A confounding factor, therefore, must have an effect and must be imbalanced between the exposure groups to be compared.

- A confounder must be associated with the disease (either as a cause or a proxy for a cause but not as an effect of the disease).
- A confounder must be associated with the exposure.
- A confounder must not be an effect of the exposure.

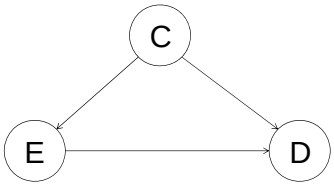
Comment: Data can only show us an association. The plausible direction of a causual effect must stem from other substantive knowledge about the phenomenon.

NTNU – Trondheim Norwegian University of Science and Technology

www.ntnu.edu

11

C is a confounder:
Adjust for C in the analysis.

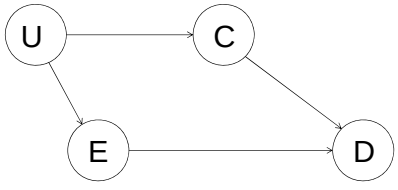


NTNU – Trondheim Norwegian University of Science and Technology

www.ntnu.edu

12

U is an umeasured confounder.
Adjusting for C removes the bias caused by U.

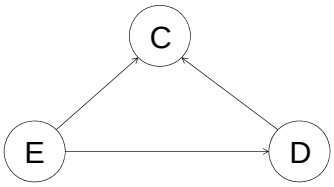


NTNU – Trondheim Norwegian University of Science and Technology

www.ntnu.edu

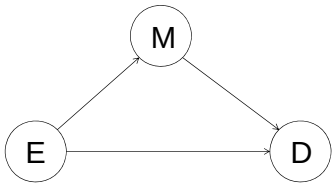
13

C is a collider:
Do not adjust for C in the analysis
– that would introduce bias.



14

M is a mediator:
Usually not appropriate to adjust for M
- then the estimated effect would be only the direct effect not mediated through M



15

How to adjust for confounders

- Separate analyses
- Stratified analysis
- Confounders as covariates in regression analysis
- Propensity score analysis

16

Limitations of separate analyses and stratified analyses

- The confounders must be
  - categorical
  - one or a few
  - few categories
- For example, 10 dichotomous confounders implies potentially 2 x 2 x…x 2 =1024 strata.

17

Limitations on number of covariates in regression analysis:

- Traditional rules of thumb: At least 10 cases per covariate (Some authors say 20 or 15 or 5).
- In logistic regression: This is the number of cases in the smallest outcome group. In survival analysis (f ex Cox regression): This is the number of events (uncensored observations).
- In logistic and Cox regression, 10 events per variable is usually sufficient (Peduzzi et al. 1996)
- In many situations 5 events per variable is sufficient (Vittinghoff & McCulloch 2007)
- The number of candidate variables must include all variables screened for association with the response (Harrell, 2001, page 61). Hence, with stepwise selection of variables, the number of candidate variables is counted, not the final list of selected variables.

18

Counting the number of covariates

- This is the number of parameters in the regression model
- Categorical covariates with k>2 categories count as k-1.
- Nonlinear effects increase the number of covariates. F ex if x_1 and x_1^2 are included, this gives 2 covariates.
- Interaction terms must be counted.

19

Stepwise selection of covariates:

Automated variable selection procedures like stepwise selection used to be very popular. Today an increasing number of analysts criticize such methods.

Rothman, K J, Greenland, S, Lash, T L: (2008) "Modern epidemiology" 3rd ed, Lippincott Williams & Wilkins, Page 419 (Chapter "Introduction to regression modelling" Section "Model searching"):
"There are several systematic, mechanical, and traditional algorithms for finding models (such as stepwise and best-subset regression) that lack logical and statistical justification and that perform poorly in theory, simulations and case studies … One serious problem is that the P-values and standard errors (SE) … will be downwardly biased, usually to a large degree."

Stepwise procedures give biased regression coefficients (the coefficients for remaining variables are too large); see Tibshirani, Journal of the Royal Statistical Society, B Series 58: 267–288, 1996).

20

Propensity score

- "In studies that do not use random allocation, this a value that indicates (separately for each subject) how likely a subject is to receive any one of the treatments being compared, given a set of covariates measured on that subject." (Day, Dictionary for clinical trials, 2nd ed, Wiley 2007)

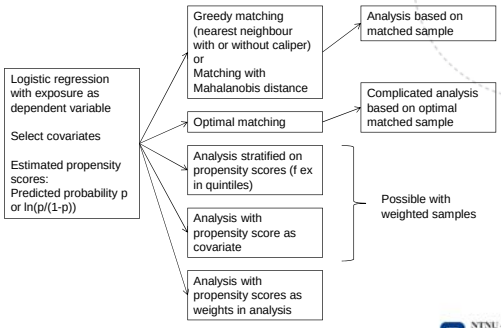
21

Propensity score

- The propensity score is the **probability** of being exposed, given the covariates. Usually modelled in logistic regression.
- Some authors (f ex Rosenbaum and Rubin) prefer using the **log odds** of this probability as the propensity score
- Propensity scores can be used only if the exposure (or treatment) variable is dichotomous (alternatively **few** categories)
- The great advantage of propensity score is its **reduction of dimensions** in matching, stratification or adjustment.
- Exposure (or treatment) assignment is considered random conditionally given the propensity score. The purpose is to mimic a randomized controlled trial (RCT).
- Propensity score methods have largely emerged from applications in economy, behavioral sciences and health science.

22

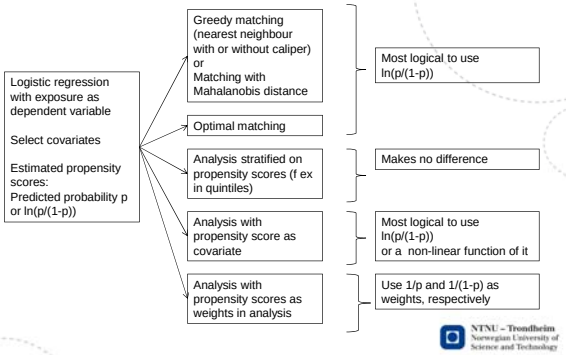
The 2 or 3 steps in propensity score analysis



Adapted from Guo and Fraser (2010) and Katz (2010)

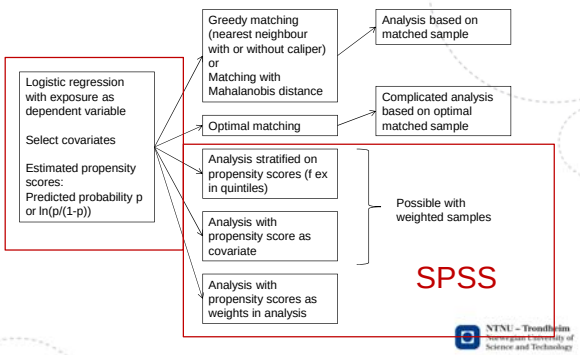
23

What to use as propensity score



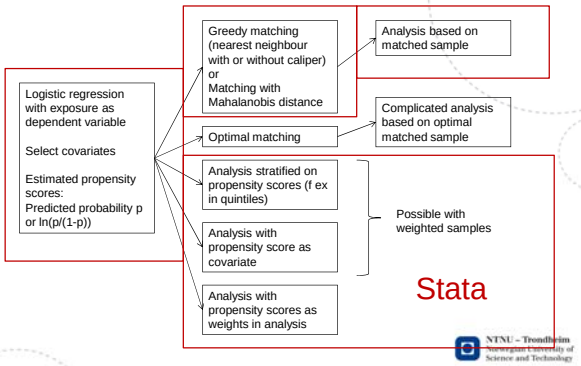
24

The 2 or 3 steps in propensity score analysis



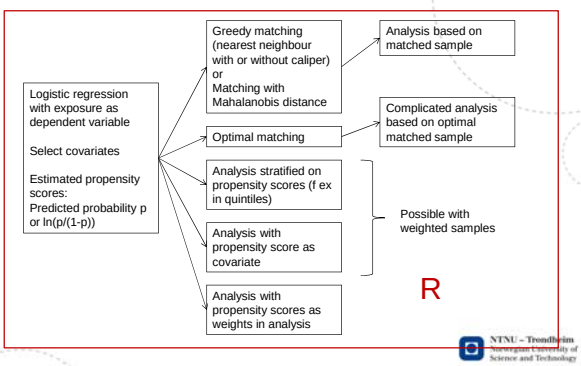
25

The 2 or 3 steps in propensity score analysis



26

The 2 or 3 steps in propensity score analysis



27

Step 1: Modelling and estimating the propensity score

- Predicting the exposure as outcome, usually logistic regression
- Include plausible predicitors of exposure, possibly with nonlinear effects and interactions.
- This is done without regard to the disease outcome. ”This means that it is possible to experiment with inclusion of different combinations of variables in the propensity score without risking biasing your model by choosing varibles based on how they affect your estimate of outcome.” (Katz, 2010, page 103)

28

Variable selection for the propensity score

- The purpose is **prediction**, not hypothesis testing. Hence, a ”rich” model with many covariates can be used.
- Rosenbaum and Rubin (1984) recommend stepwise selection with a low threshold for significance, such as |t|>1.5 (p<appr. 0.15)
- Comment 1: Need not exclude variables using stepwise if the exposure (and non-exposure) groups are large compared to the number of variables
- Comment 2: Stepwise selection is controversial and not recommended in models for hypothesis testing (step 2-3). Less controversial(?) in propensity score modelling (step 1)

29

Variable selection for propensity score

- Rosenbaum and Rubin (1984, 1985) use high-order polynomial terms and/or cross-product interaction terms.
- Comment: Fractional polynomials (Royston and Altman, 1994) work better than high-order polymonials in regression modelling. See f.ex Fagerland, Eide, Laake: Chapter 4: ”Linear regression” in Veierød, Lydersen, Laake (eds): ”Medical statistics in clinical an epidemiological research”, Gyldendal Akademisk 2012.

30

Propensity score model check

- Check that the potential confounders (included variables) are evenly balanced in matched data.
- If not well balanced, try including higher order (or fractional polynomial) terms for that variable, or interactions
- Check balance again

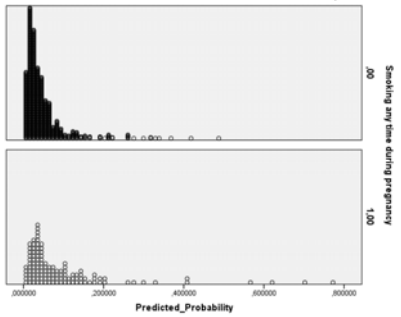
31

Ellis et al (2012) Propensity score

- Mother's age, SES, and antisocial personality traits were included.
- In addition, stepwise selection included 11 of the 24 potential confounders:

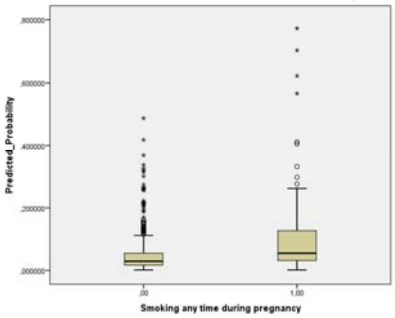
- borderline personality traits
- parental alcohol use
- parental anxiety
- alcohol use during pregnancy
- depression during pregnancy
- planned pregnancy
- mothers' feelings in the first month after birth
- parent ever been arrested
- parent ever been indicted by police
- parental ability to pay family expenses
- parental admission to a mental health institution

32

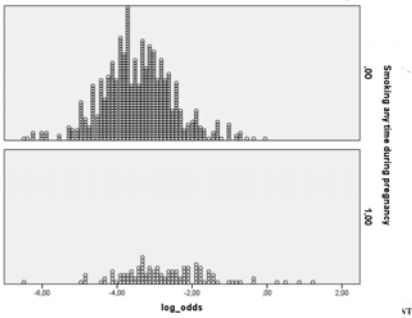


SD = 0.073, 0.25SD=0.018

33

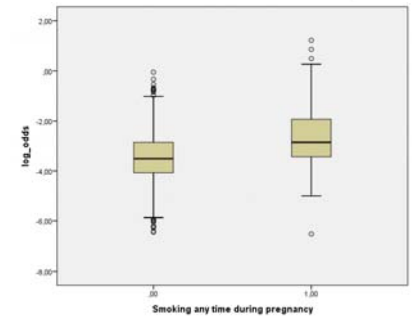


34



SD = 1.03, 0.25SD=0.26

35



36

Table 2 Propensity score quintiles and smoking during pregnancy

Propensity score quintile	Smoking during pregnancy, n (%)		Total
	No	Yes	
1	158 (95.8)	7 (4.2)	165
2	157 (95.2)	8 (4.8)	165
3	149 (90.9)	15 (9.1)	164
4	134 (80.2)	33 (19.8)	167
5	108 (66.3)	55 (33.7)	163
Total	706 (85.7)	118 (14.3)	824

37

Step 2: Propensity score greedy matching

- The propensity score P_j for subject number j is the estimated probability p of being exposed.
- Alternatively, the propensity score is the log odds ln(p/(1-p)). Rosenbaum and Rubin (for some reason) use ln((1-p)/p).
- Nearest neighbour: For an exposed subject i, choose the non-exposed subject j (without replacement) which minimizes $|P_i - P_j|$.
- Nearest neighbour with caliper: Include this matched pair only if $|P_i - P_j| < \varepsilon$. Recommended $\varepsilon=0.25\sigma_p$.
- 1-to-1 nearest neighbour with caliper is a common practice
- 1-to-n nearest neighbour with caliper

38

Alternative: Mahalanobis distance greedy matching

- For two subjects with covariate vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, the Mahalanobis distance is $(\mathbf{x}_i - \mathbf{x}_j)^T\Sigma_{\mathbf{x}}^{-1}(\mathbf{x}_i - \mathbf{x}_j)$.
- Mahalanobis without propensity score
- Mahalanobis with propensity score added (to $\mathbf{x}$)
- Mahalanobis within calipers defined by propensity score (need your own programming)

39

Post-matching analysis (step 3)

- Analyze as if data were from an RCT
- Include the matching variable in the analysis
- Analysis methods may be one or more of these:
  - Linear regression
  - Logistic regression (or other generalized linear model)
  - Survival analysis
  - Structural equation modelling (SEM)
  - Mixed model (for longitudinal or hierachically clustered data)
  - Generalized estimating equations (GEE)

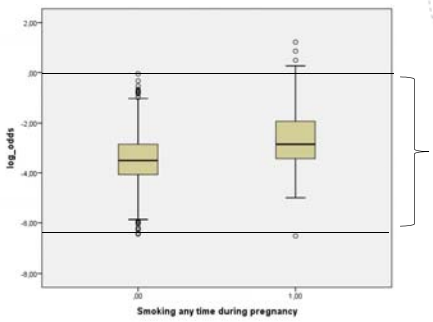
40

Limitations of greedy matching

- Dilemma between incomplete matching and inaccurate matching.
- Not all exposed and non-exposed are included in the analysis
- Needs a sizeable common-support region

41

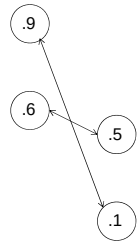
Limitations of greedy matching



Cases outside the common support region are eliminated

42

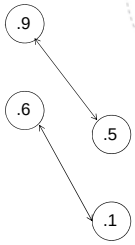
Greedy matching:



Exposed Non-exposed

Total distance:
|.5-.6|+|.1-.9|=.9

Optimal matching:



Exposed Non-exposed

Total distance:
|.5-.9|+|.1-.6|=.7

43

Optimal matching

- Pair matching: Each exposed is matched to a single non-exposed
- Variable matching: Each exposed is matched to, for example, at least one and at most four non-exposed
- Full matching: Each exposed is matched to at least one non-exposed, and each non-exposed is matched to at least one exposed.

May analyze as if from an RCT

Need specialized (complicated) analysis

44

Analysis stratified on propensity scores

- Makes no difference if you use p or ln(p/(1-p))
- Creating five strata (by quintiles of the propensity score) removes about 90% of the bias in unadjusted analysis (Rubin, 1997)
- Allows for non-linear effects of the propensity score on the outcome
- Five strata create four parameters if included as a categorical covariate.
- Alternatively, one may use Mantel-Haenszel methods if the outcome is dichotomous.

45

Propensity score as covariate

- More logical to use the log odds, ln(p/(1-p)), than to use p.
- Introduces only one covariate in the model
- May be sensible to consider a non-linear function of the propensity score

46

Propensity scores as weights in the analysis

- For exposed subjects, use 1/p as probability weight
- For non-exposed, use 1/(1-p) as weight
- Use ”Complex samples” in SPSS or the option pweight in Stata (easier).
- The subject may be thought of as representing 1/p (or 1/(1-p)) subjects.
- If some p are near 0 or 1, these weights may be extreme and unrealistic. Guo suggests removing those subjects from the analysis.
- Controversial procedure if the sample is (originally) weighted like in Ellis et al (2012).

47

Methods for using propensity scores

Method	Advantages	Disadvantages
Greedy matching (with or without caliper)	Obtain comparable groups	Decreased sample size
Optimal matching	Obtain comparable groups	Complicated analysis
Stratification (for ex in quintiles)	Allows inclusion of subjects otherwise lost due to no close matches	Residual bias
Propensity score as covariate in analysis	Allows inclusion of subjects otherwise lost due to no close matches	Residual bias
Propensity scores to weight observations	Allows inclusion of subjects otherwise lost due to no close matches. May be less subject to misspecification of analysis model	Propensity scores near 0 or 1 may create problems

Adapted from Katz (2012) Table 7.3 page 110

48

Choice of method

- ”Use matching by propensity score first and then consider stratifying by propensity score or including the propensity score as a covariate to improve generalizability. Different methods of using propensity scores should lead to similar results.” (Katz, 2012, page 115)

49

Choice between stratification and propensisty score as covariate:

- ”Between the two methods, I would say that stratification is better. At the lowest and highest quintile, the treated and nontreated groups are generally balanced on propensity scores. Rosenbaum and Rubin's (1983) Corollary 4.2 is a proof of this property. Using estimated propensity score as an independent variable is valid, only if you assume that the covariates affecting selection are the same factors affecting outcome. In real data, this may not be the case. If the original data suffer from the problem of endogeneity, including a propensity score in the regression may not remove bias, because the residual term (i.e., the unmeasured variance) may still be correlated with the treatment variable.” (Personal communication from Shenyang Guo to Stian Lydersen, November 2012)

50

Table 3 Odds ratio (OR estimate, CI and p values) for psychiatric disorders, for children exposed to smoking any time during pregnancy (Ellis et al 2012)

	ADHD (n=34)	ODD (n=57)	ADHD and ODD (n=13)
Unadjusted	3.25 (2.08–5.09) p<0.001	3.12 (2.30–4.24) p<0.001	3.67 (1.82–7.40) p<0.001
Adjusted for propensity score stratified in quintiles	2.59 (1.50–4.34) p<0.001	2.69 (1.84–3.91) p<0.001	3.69 (1.68–8.14) p<0.001
Adjusted for propensisty score (probability) as covariate	2.17 (1.30–3.61) p = 0.003	2.46 (1.66–3.63) p<0.001	2.68 (1.84–3.91) p<0.001

51

Hindsight: The analysis in Ellis et al (2012).

- Stepwise selection of covariates in the propensity score modelling was OK according to Guo.
- Could have included more variables in the propensity score model. Could have used p<0.15 instead of p<0.05.
- Should have checked balance of covariates within propensity score quintiles.
- When using the propensity score as covariate: Would have been more logical to use log odds instead of p.
- OK that we refrained from using propensity score weighting in a weighted sample. That would have been controversial.